Subtle adjustments of the glucose-6-phosphate dehydrogenase (G6PD) mutation database and reference sequence.

Stéphane Mazières, Florence Petit, Jean-Michel Dugoujon, Xavier Iriart, Antoine Berry, Bernard Carme, Mathieu Nacher, Pascal Bailly, Jacques Chiaroni

Letter to the Editor

Subtle adjustments of the Glucose-6-phosphate dehydrogenase (G6PD) mutations database and reference sequence.

To the Editor:

Reference sequences and mutation databases are essential for the development of molecular-based methods in human genetics. Lately, Minucci et al. [1] revised the glucose-6-phosphate dehydrogensase (G6PD) reference material from 131 bibliographic references, three previous databases, and the genomic reference sequence (GenBank accession number X55448.1). Deficiency in G6PD is the most common enzymatic insufficiency in human populations and clinical manifestations range from mild to severe: neonatal jaundice, acute hemolysis and chronic non-spherocytic hemolytic anemia prompted by infections or sudden onset of oxidative agents from for instance the ingestion of fava beans (*Vicia faba*) or anti-malarial treatments [2]. Given the clinical consequences, numerous studies and international projects are being carried out about the mechanisms leading to the numerous G6PD enzymatic variants, and in this context, the updated G6PD database provided in [1] is an precious substratum for lab protocols. Nonetheless, while setting our strategy from this material, we encountered two main hurdles that we would like to address.

In order to accurately design our sequencing primers, we started by manually relocating each of the 187 listed substitutions onto X55448.1 by the addition of 3350 to their respective genomic DNA nucleotide position (gDNA np) [1]. We then validated each targeted substitution according to the recognition enzyme site it should create or abolish. Occasionally, we have observed a mismatch between some nucleotides listed in [1] and the G6PD reference sequence X55448.1. For instance, the Rignano mutation involving a G or A was expected at np13337 in X55448.1. However, np13337 is a C while the preceding nucleotide is a G. Relocation of the substitutions revealed that the *G6PD* mutation database mismatched with X55448.1 by one base at exons 3, 4 and 5. We presume that this numbering shift originated from the update of the former X55448 involving a deletion of one T before exon 3 at np8850 [1]. In order to asses our suspicion, we then ran a multiple sequence alignment of several *G6PD* reference sequences using BioEdit version 7.0.5.3 [3]: ultimate (X55448.1) , former

(X55448), *Homo sapiens* RefSeqGene (NG_009015.1) and coding DNA (X03674.1). Sequence alignment confirmed that the entire block encompassed between np8850 and np14575 of X55448.1 is shifted one-base toward the left. This "sliding-effect" is then counter-balanced from intron 5 by the addition of a C at np14575; another major update of the former X55448 [1]. Supplementary Table 1 lists the 37 substitutions affected by the incident among which 22 are unambiguous.

Concomitantly, the multiple sequence alignment showed that X55448.1 carried an A at np13503 and a G at np14226 (Table 1), characteristic of the $A^{-(202)}$ variant [4]. This is problematic since all G6PD enzymatic variants are given according to the wild-type B variant, and the reference sequence should mirror it (Figure 1). We sequenced exons 4 and 5 in two male samples free from G6PD deficiency using the primers previously published [5]. Capillary electrophoresis was carried out on an automated fluorescence-based ABI PRISM ® 3130 XL genetic analyzer (Applied Biosystem, Forster City, CA, USA) according to the manufacturer's protocol. Sequencing and alignment of our two samples confirm the non-ancestral allelic state of the G6PD reference sequence X55448.1 of the NCBI BioSystems database [6].

Additionally, we found four minor misprints: the Nashville-Anaheim-Portici and Georgia variants were spotted at gDNA np13447 and 13560 while they are actually at gDNA np13477 and 13583. GDNA coordinate of the $A^{-(680)}$ mutation was erroneously written 120228, and the Acrokorinthos variant has been labeled in exon 12 while it is in exon 5 [7] (Supplementary Table1).

In conclusion, we authenticated that the nucleotide substitutions of exons 3, 4, and 5 from the latest mutation database have been erroneously numbered by +1, and the updated *G6PD* reference sequence appears to be non-wild-type. Four minor misprints have also been spotted. Herein we provide the slight modifications to favor the standardization of the *G6PD* reference material: firstly subtracting 1 to the gDNA coordinates of the nucleotide variations of exons 3, 4, and 5; and secondly, reversing two punctual polymorphisms to turn over the reference sequence to the wild-type allele status. *G6PD* reference sequence and mutation database of [1] remain essential to the identification of enzymatic variants in areas of malaria exposure from molecular-based attempts.

Acknowledgement:

This work has benefited from an *Investissement d'Avenir* grant of the ANR (CEBA: ANR-10-LABX-0025).

References

[1] A. Minucci, K. Moradkhani, M.J. Hwang, C. Zuppi, B. Giardina, and E. Capoluongo, Glucose-6-phosphate dehydrogenase (G6PD) mutations database: review of the "old" and update of the new mutations. Blood Cells Mol Dis 48 (2012) 154-65.

[2] E. Beutler, Glucose-6-phosphate dehydrogenase deficiency: a historical perspective. Blood 111 (2008) 16-24.

[3] T.A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41 (1999) 95-98.

[4] T.J. Vulliamy, M. D'Urso, G. Battistuzzi, M. Estrada, N.S. Foulkes, G. Martini, V. Calabro, V. Poggi, R. Giordano, M. Town, and et al., Diverse point mutations in the human glucose-6-phosphate dehydrogenase gene cause enzyme deficiency and mild or severe hemolytic anemia. Proc Natl Acad Sci U S A 85 (1988) 5171-5.

[5] A. Minucci, L. Gentile, C. Zuppi, B. Giardina, and E. Capoluongo, Rapid and simple identification of the commonest glucose-6-phosphate dehydrogenase (G6PD) Italian mutations: from DNA extraction to genotyping. Clin Chim Acta 413 (2012) 1018-9.

[6] L.Y. Geer, A. Marchler-Bauer, R.C. Geer, L. Han, J. He, S. He, C. Liu, W. Shi, and S.H. Bryant, The NCBI BioSystems database. Nucleic Acids Res 38 (2010) D492-6.

[7] A. Drousiotou, E.H. Touma, N. Andreou, J. Loiselet, M. Angastiniotis, B.C. Verrelli, and S.A. Tishkoff, Molecular characterization of G6PD deficiency in Cyprus. Blood Cells Mol Dis 33 (2004) 25-30.

Stéphane Mazières

Florence Petit

Aix Marseille Université, CNRS, EFS-AM, ADES UMR7268, 13344, Marseille, France.

Jean-Michel Dugoujon

Université Paul Sabatier, CNRS, AMIS UMR5288, 31073, Toulouse, France

Xavier Iriart

Antoine Berry

Service de Parasitologie-Mycologie, Centre Hospitalier Universitaire de Toulouse / INSERM

UMR1043/CNRS UMR5282/Université de Toulouse UPS, Centre de Physiopathiologie de Toulouse

Purpan (CPTP), 31300, Toulouse, France

Bernard Carme,[5]

Mathieu Nacher,[5]

CIC-EC Antilles Guyane CIE 802 Inserm, Centre Hospitalier Andrée Rosemon

Equipe EPaT EA 3593, Université des Antilles et de la Guyane, Cayenne, Guyane française.

Pascal Bailly

Jacques Chiaroni[*]

[3]*Établissement Français du Sang Alpes Méditerranée, 13005, Marseille, France*

Email address: jacques.chiaroni@efs.sante.fr

*Corresponding author. Fax: +33 (0)4 9118 9350

Table 1 Nucleotide variations in a set of *G6PD* reference sequences and samples. Coding (cDNA) and genomic (gDNA) coordinates correspond to the locations on X03674.1 and X55448.1 respectively.

Sequence	GenBank Accession number	202G->A (cDNA np672, gDNA np13503)	376A->G (cDNA np846, gDNA np14226)	Allele
H. sapiens G6PD RefSeqGene	NG_009015.1	G	A	B
cDNA	X03674.1	.	.	B
Our samples	N/A	.	.	B
gDNA *H. sapiens G6PD* gene	X55448.1	A	G	$A^{-(202)}$

Supplementary Table 1 List of the 37 *G6PD* mutations whose gDNA numbering should be modified by minus 1 as inferred from X55448.1. The four misprints (underlined) were detected after comparison with X03674.1 and recognition enzyme sites [1].

Substitution	Exon	Mutation name	gDNA np	Observed nucleotide	Accurate nucleotide	Rectified gDNA np
1.One-base shifting						
130G>A	3	Rignano	9987	13337C	13336G	9986
131C->G	3	Orissa	9988	13338C	13337C	9987
143T->C	3	Aures	10000	13350C	13349T	9999
148C->T	3	Kambos	10005	13355C	13354C	10004
159G->C	4	Kozukata	10111	13461T	13460G	10110
169C->T	4	Kamogawa	10121	13471G	13470C	10120
170G>A	4	Palestrina	10122	13472G	13471G	10121
172G->A	4	Metaponto	10124	13474A	13473G	10123
179T>C	4	Costanzo	10131	13481T	13480T	10130
180_182del TCT	4	Amsterdam	10132–34	13482-84CTG	13481-83TCT	10131–33
185C->A	4	Amazonia-Musashino	10137	13487C	13486C	10136
196T->A	4	Songklanagarind	10148	13498T	13497T	10147
202G->A	4	Asahi, Hechi, $A^{-(202)}$, no name	10154	13504T	13503G	10153

Supplementary table 1. Continued

Substitution	Exon	Mutation name	gDNA np	Observed nucleotide	Accurate nucleotide	Rectified gDNA np
1.One-base shifting						
208T->C	4	Namouru	10160	13510A	13509T	10159
209A->G	4	Murcia Oristano	10161	13511T	13510A	10160
224T->C	4	Swansea	10176	13526C	13525T	10175
241C->T	4	Ube, Konan	10193	13543G	13542C	10192
242G->A	4	Lagosanto	10194	13544C	13543G	10193
274C->T	5	Guangzhou	10775	14125C	14124C	10774
281_283del AGA	4	Urayasu	10782–84	14132-34GAA	14131-33AGA	10781–10783
317C->G	5	Vancouver	10818	14168C	14167C	10817
323T->A	5	Hammersmith	10824	14174G	14173T	10823
337G->A	5	São Borja	10838	14188A	14187G	10837
352T>C	5	Bao Loc	10853	14203A	14202T	10852
375G->T	5	Crispim	10876	14226G	14225G	10875
376A->G	5	A, Acrokorinthos, Santamaria, $A^{-(680)}$, Ananindeua, $A^{-(968)}$, Betica, Selma, Guantanamo	10877	14227A	14226A	10876

Supplementary table 1. Continued

Substitution	Exon	Mutation name	gDNA np	Observed nucleotide	Accurate nucleotide	Rectified gDNA np
1.One-base shifting						
379G->T	5	Crispim	10880	14230C	14229G	10879
383T->C	5	Vanua Lava-Crispim-Salerno Pyrgos	10884	14234C	14233T	10883
384C>T	5	Crispim	10885	14235C	14234C	10884
392G->T	5	Quing Yan	10893	14243G	14242G	10892
404A->C	5	Cairo	10905	14255C	14254A	10904
406C->T	5	Valladolid	10907	14257G	14256C	10906
409C->T	5	Belem	10910	14260T	14259C	10909
442G->A	5	Liuzhou	10943	14293A	14292G	10942
466G->A	5	Ilesha	10967	14317A	14316G	10966
473G>A	5	Shenzen	10974	14324C	14323G	10973
477G>C	5	Gond	10978	14328A	14327G	10977
2.Misprints						
1178G->A		Nashville, Anaheim, Portici	13447	1648G	16827G	13477

Supplementary table 1. Continued

Substitution	Exon	Mutation name	gDNA np	Observed nucleotide	Accurate nucleotide	Rectified gDNA np
2.Misprints						
1284C->A		Georgia	13560	1754A	16933C	13583
		$A^{-(680)}$	120228			12028
	5	Acrokorinthos				

Figure 1. Extract of X55448.1 with the main modifications turning over the sequence from the $A^{-(202)}$ to the B variant at np13503 and np14226 (open boxes)